



The Effect of Outliers on Measures of Centre

How to Get a Random Sample from CensusAtSchool

Go to the CensusAtSchool [Random Sampler](#) to download a sample.X

Reference year: *(select year)* **Sample size:** *10 students*

Select questions: *Foot length*

Location: *Select location* **Year level:** *(select a range of year levels)*

To protect privacy there is a rule built into the sampler that the requested sample size cannot exceed 10% entered.

Task One:

The mean, median and mode are also called the measures of centre of the data.

Statisticians calculate the mean, median and mode of data so that they can get a feel for what is 'average' in a set of data. These measures of centre can be easier to use when describing data rather than referring to each individual piece of data.

For example, there are times that we might want to compare ourselves to the average person for our age..... Are we fitter, taller, richer or smarter?

1. Write each measure of centre next to its definition below.
2. In your CensusAtSchool random sample find the question about foot length and calculate the mean, median and mode of the ten numbers.

Definition	Measure of centre (Mean, median or mode)	Calculation (Show working)
The middle value in an ordered data set		
The most frequently occurring value		



Add all the values then divide by the number of values		
--	--	--

Task Two: When do we use each measure of centre?

Categorical Data

The mode is the most frequently occurring and is suitable to use for categorical or non-numerical data.

Numerical Data

Both the mean and the median are used for data that is in numerical form.

You are going to investigate when it is more appropriate to use the median rather than the mean as the typical value when numerical data is collected.

The Guinness Book of Records shows that the longest foot length is 68.58 cm. Replace the longest foot length in your random sample with this length.

3. Recalculate, showing your working:
 - a) the mean
 - b) the median
4. Imagine someone made a mistake and recorded their foot length as 10cm instead of 30cm. This time, replace the shortest foot length in your original sample with this length and recalculate:
 - a) mean
 - b) median
5. Statisticians refer to the median as more resistant to extreme values than the mean. Explain why this is.
6. In general when would it be more appropriate to use the median rather than the mean to describe a typical value for a data set?
7. Why is it more appropriate to use the mode rather than mean or median as the measure of centre for type of internet connection in the Census at School data?



8. Annual house prices for an area are usually reported as the median value. Why do you think this is?